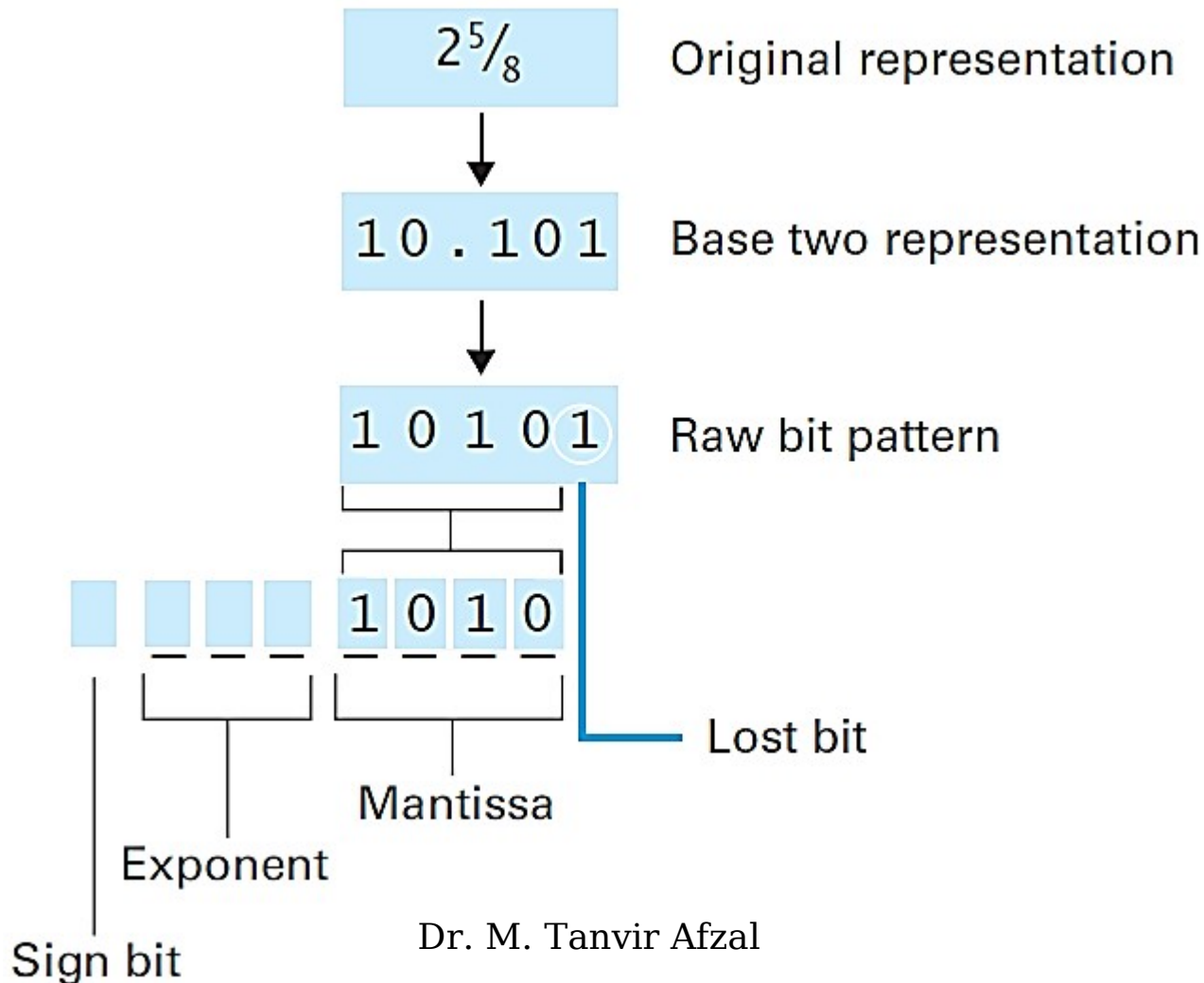


Data Storage

Truncation Errors in Floating Point Notation

Truncation Errors



Handling Truncation Errors

Primitive Methods

- ✓ Use more bits – today's computer use 32 bits
- ✓ Change the units
- ✓ Mathematics – Numerical Analysis

Handling Truncation Errors

Intelligent Processing

- ✓ Lets suppose we want to store
- ✓ $2^{-1/2} + 1/8 + 1/8$
- ✓ If we add $2^{-1/2}$ to $1/8$ we ends up with $2^{-5/8}$ which is 10.101 which can not be stored in 4 bit mantissa.
- ✓ The result 10.10 would be $2^{-1/2}$ this means the $1/8$ is truncated

Handling Truncation Errors

Intelligent Processing

- ✓ Lets add first $1/8$ to $1/8$ which is $1/4$ or $.01$, can be stored and result would be 00111000
- ✓ Now add this to $2^{-1/2}$ now we got $2^{-3/4} = 01101011$ which is accurate.
- ✓ Order is important, adding small quantities together first might give significant quantity to be added to large quantity.

Summary

Truncation errors in Floating points

- ✓ Why truncation errors occurs with examples.
- ✓ Primitive ways to deal
- ✓ Intelligent processing – the order of addition